Application example

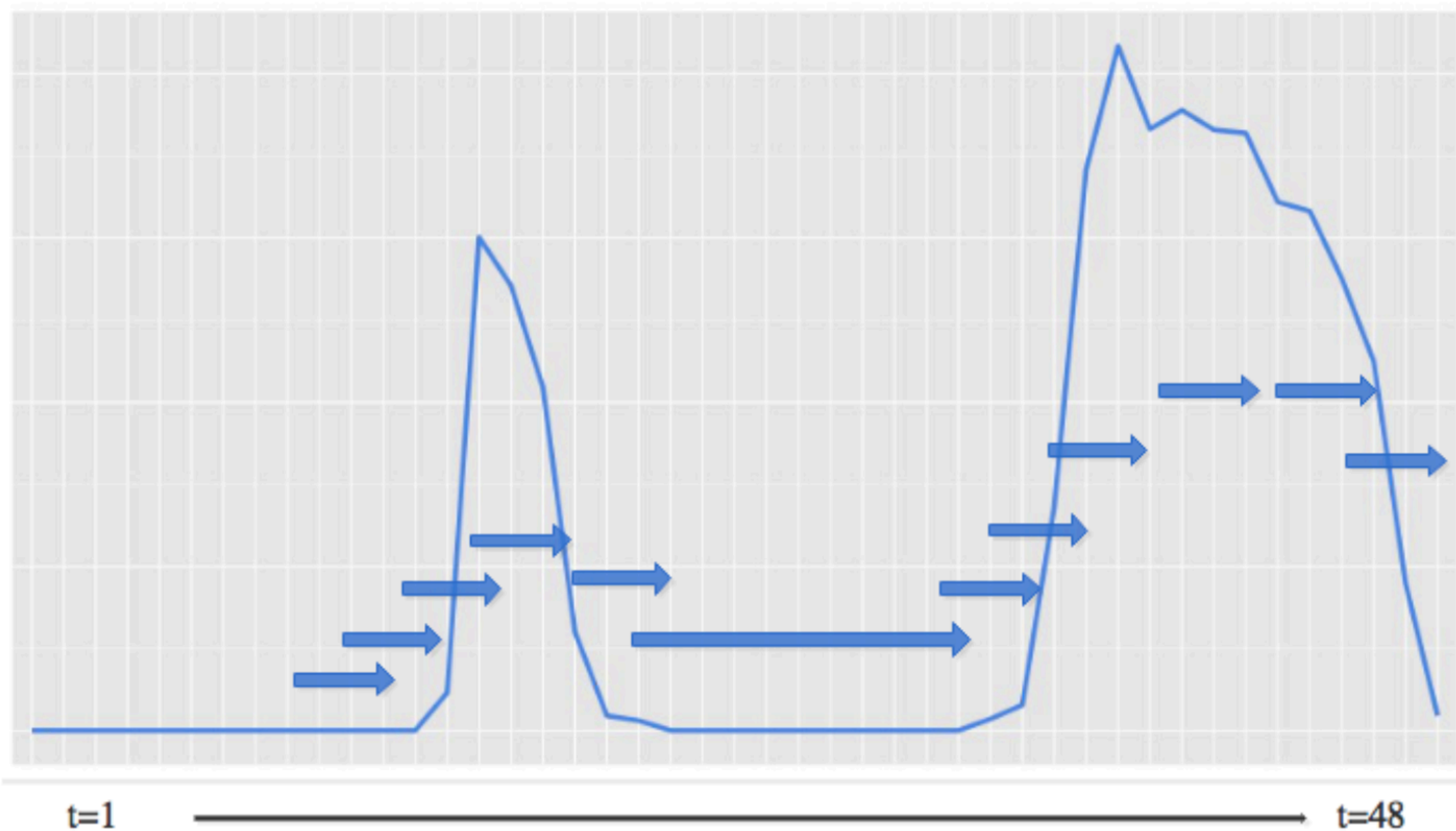
Estimating the Structure of the Non-Stationary Spatiotemporal Profiles for Behavior Prediction

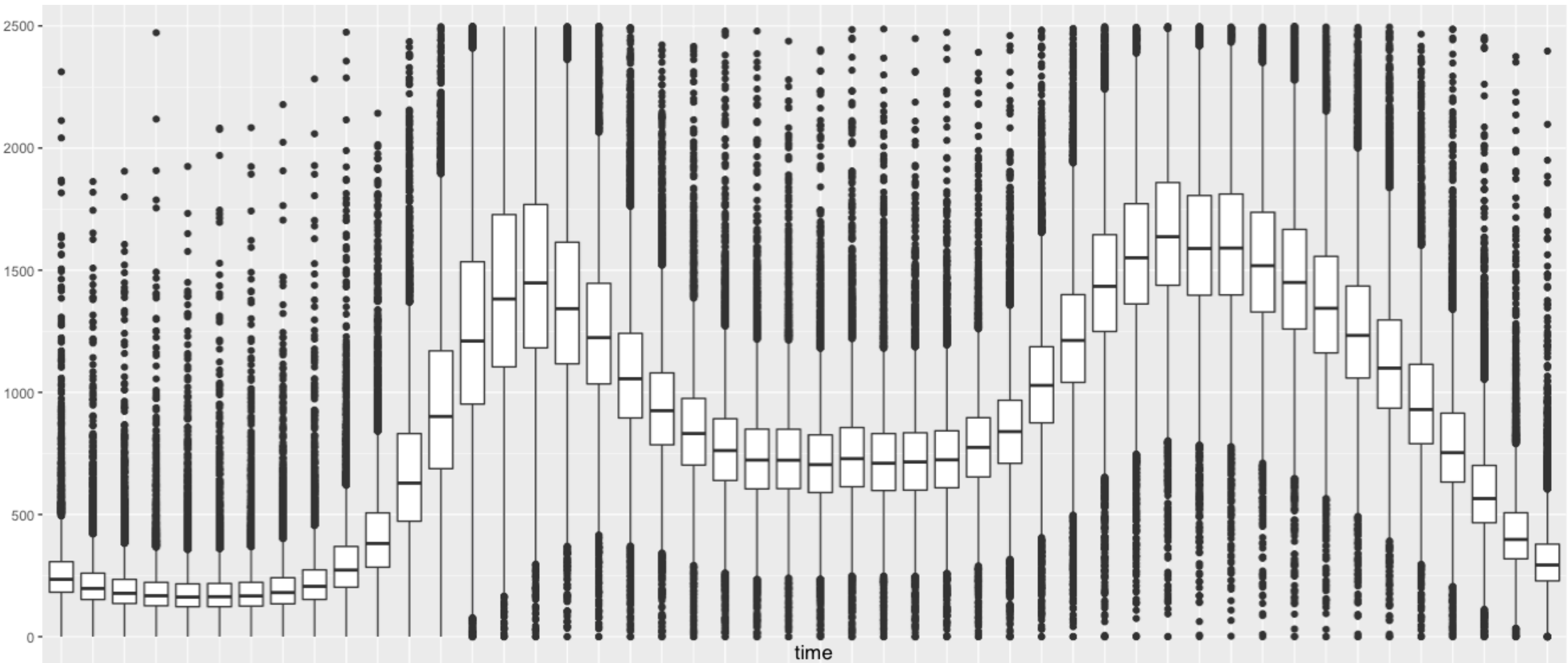
With Anastasia Ushakova (UCL, Consumer Data Research Centre)

Policy problem

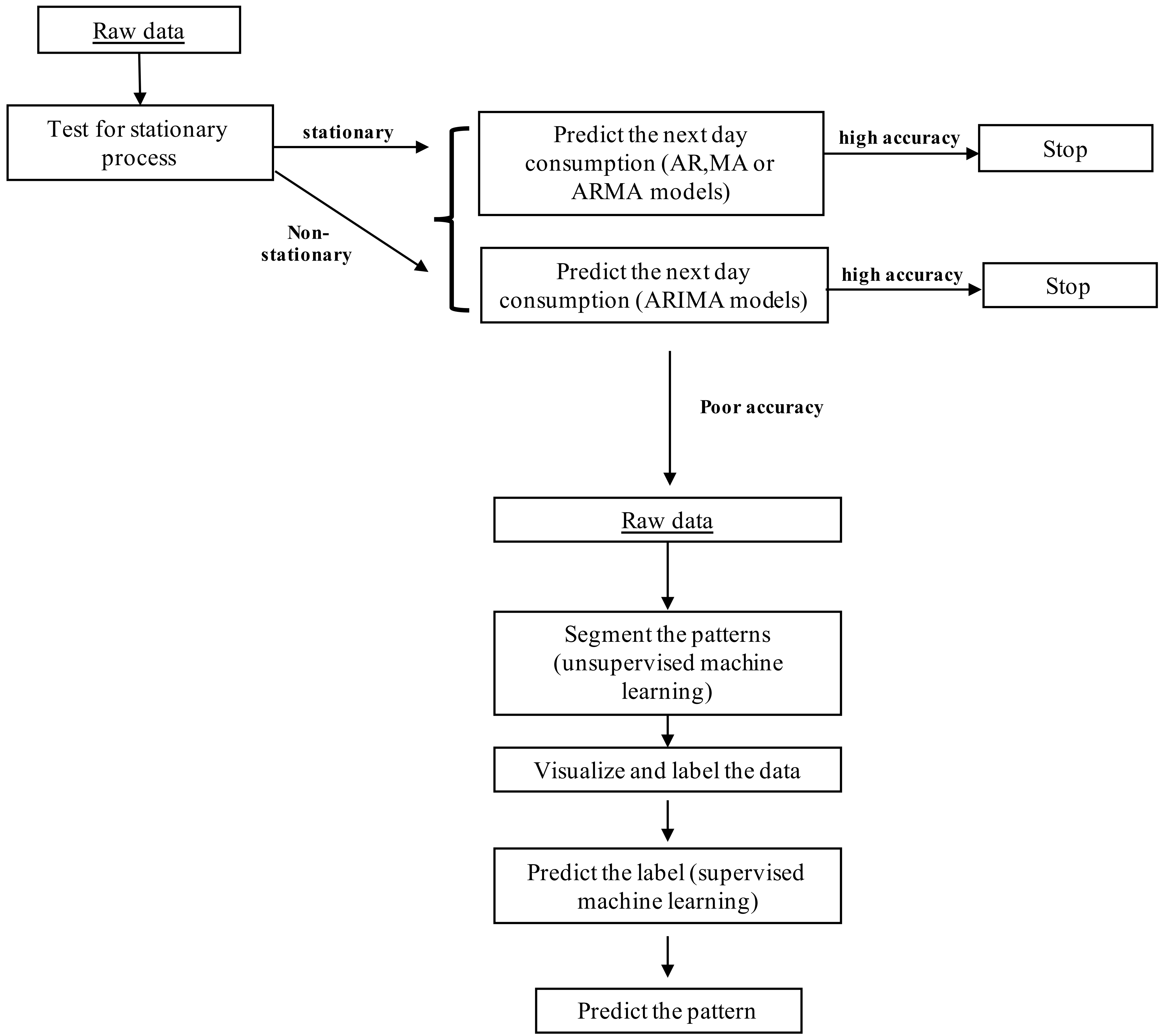
- Energy companies and identification of fuel poor.
- Energy Company Obligations (ECO) and the Green Deal.
- Conventional methods expensive and not agile.
- Consumption data available from smart meters.

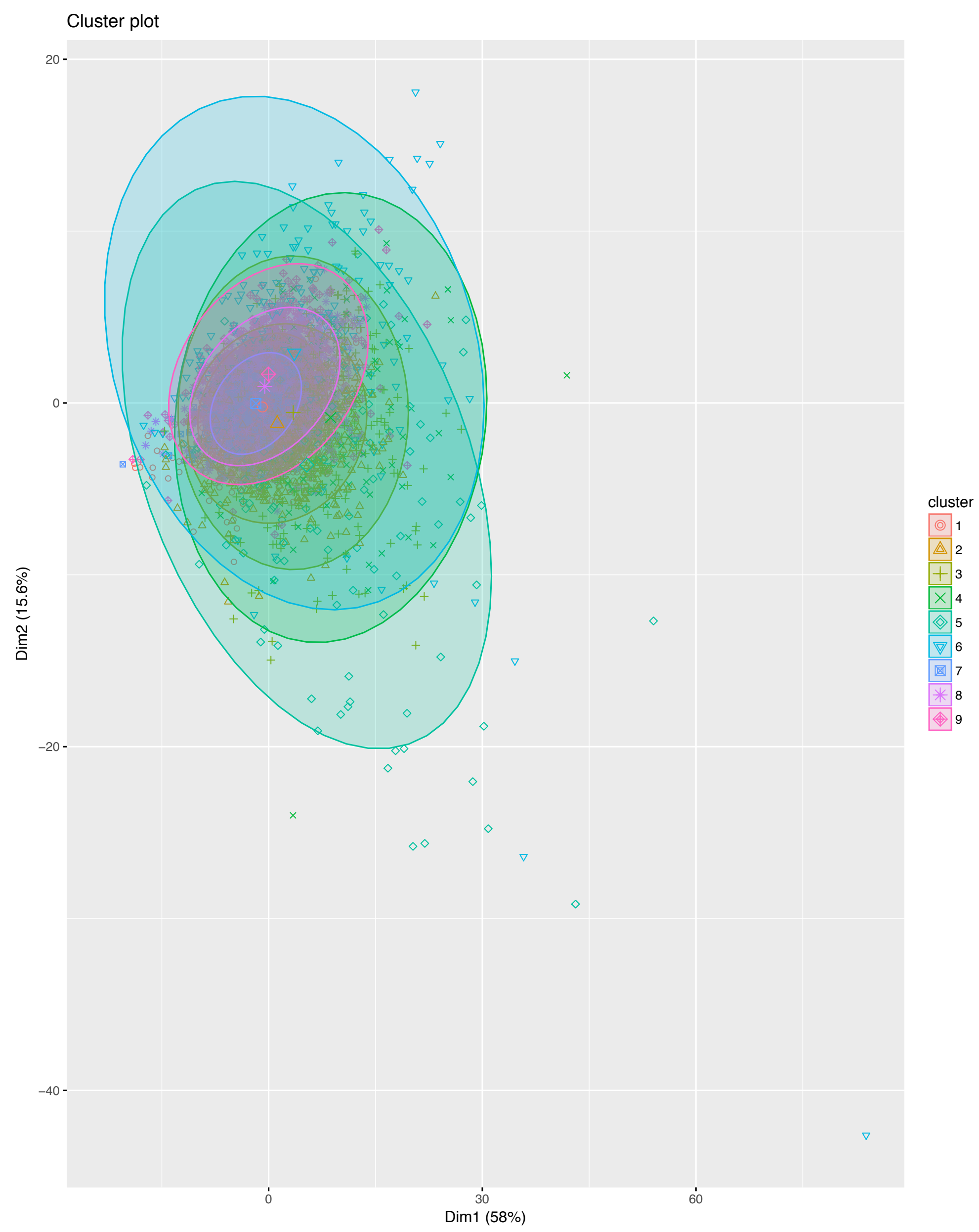
Data	Overall	Aggregated Sample	Dissagregated Sample
Unique identifiers	489,000	8,171	15100
Days	365	365	365
Daily readings)	48	48	48
Total observations	8,567,280,000	143,155,920	19,272,000


t=1
t=48



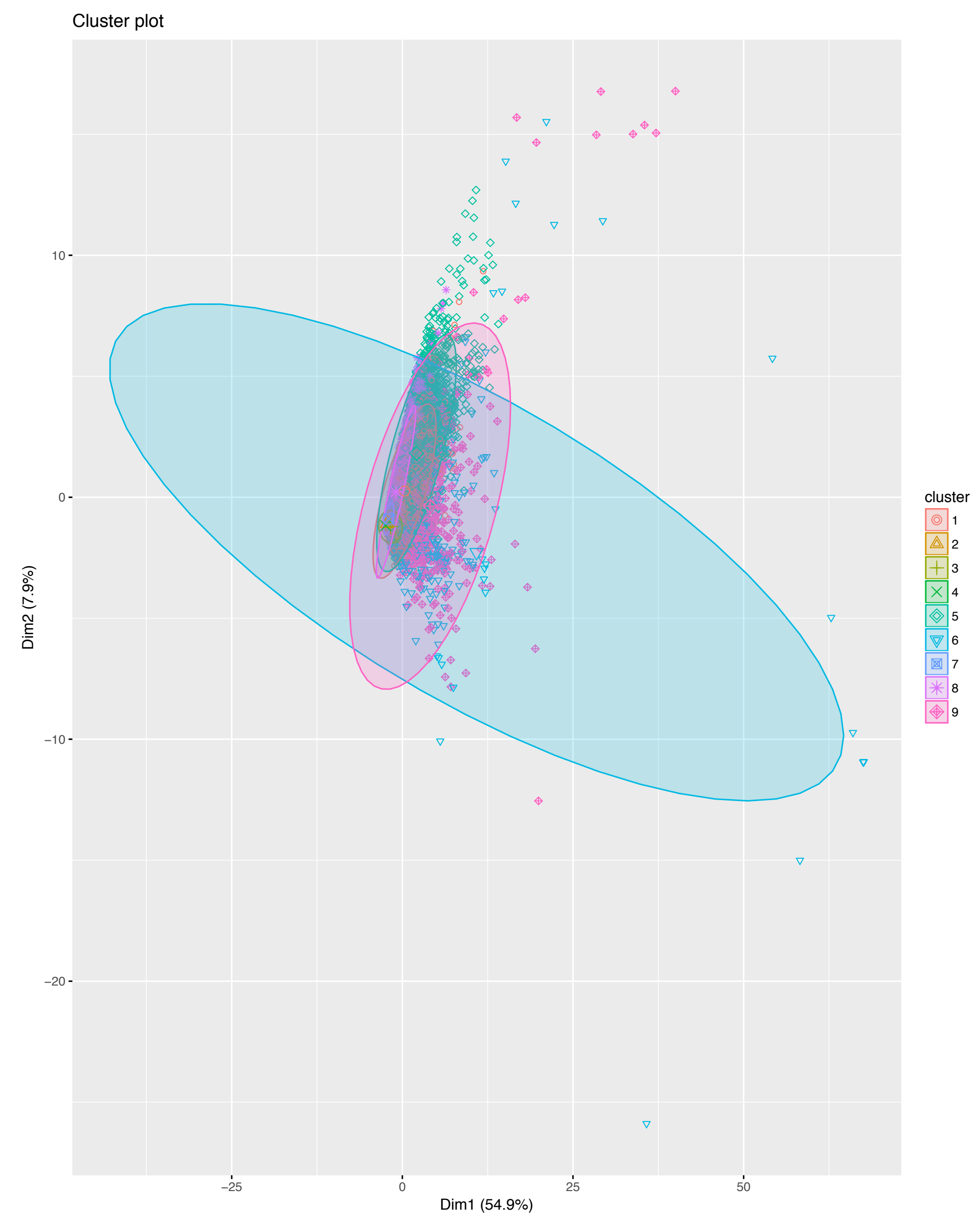
2500
2000
1500
1000
500
0
time


Raw data
Test for stationary process
stationary
Non-stationary
Predict the next day consumption (AR,MA or ARMA models)
high accuracy
Stop
Predict the next day consumption (ARIMA models)
high accuracy
Stop
Poor accuracy
Raw data
Segment the patterns (unsupervised machine learning)
Visualize and label the data
Predict the label (supervised machine learning)
Predict the pattern



Gaussian Mixture Models clustering

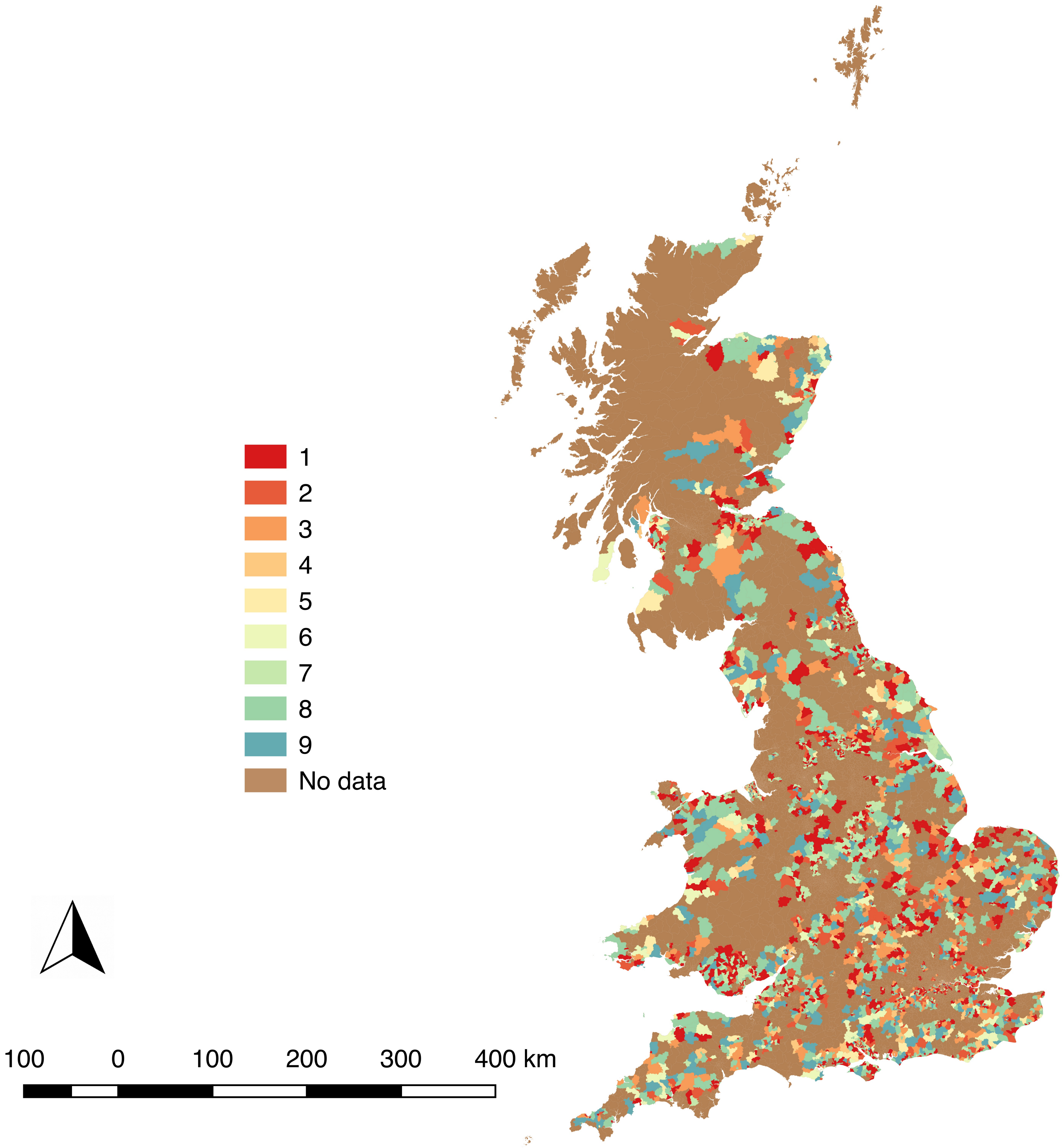
Aggregated level



Gaussian Mixture Models clustering

Disaggregated level

Segment	% of total sample (Aggregated patterns)	% of total sample (Diaggregated patterns)
1	24.0%	15.7%
2	10.6%	14.2%
3	5.3%	1.4%
4	0.9%	5.9%
5	1.9%	20.0%
6	21.9%	3.4%
7	15.5%	13.6%
8	14.4%	22.5%
9	5.5%	3.4%



1
2
3
4
5
6
7
8
9
No data
100 0 100 200 300 400 km

Aggregated Results

KNN	1	2	3	4	5	6	7	8	9
1	**13.79%**	3.42%	1.67%	0.00%	0.00%	0.00%	26.85%	5.08%	2.04%
2	15.17%	**23.29%**	0.00%	0.00%	0.00%	0.00%	9.34%	11.02%	0.00%
3	11.03%	21.23%	**6.67%**	9.52%	3.45%	10.53%	6.23%	13.56%	10.20%
4	4.83%	13.01%	40.00%	**47.62%**	34.48%	21.05%	0.00%	7.63%	16.33%
5	11.03%	14.38%	11.67%	23.81%	**41.38%**	21.05%	5.06%	8.47%	10.20%
6	6.21%	7.53%	26.67%	19.05%	20.69%	**47.37%**	1.56%	11.02%	42.86%
7	8.28%	1.37%	0.00%	0.00%	0.00%	0.00%	**30.35%**	0.85%	0.00%
8	16.55%	5.48%	1.67%	0.00%	0.00%	0.00%	14.79%	**17.80%**	2.04%
9	13.10%	10.27%	11.67%	0.00%	0.00%	0.00%	5.84%	24.58%	**16.33%**

GBM	1	2	3	4	5	6	7	8	9
1	**27.03%**	11.54%	1.52%	1.39%	1.32%	1.19%	24.79%	15.38%	2.15%
2	**12.61%**	**36.54%**	21.21%	4.17%	5.26%	0.00%	5.98%	8.65%	4.30%
3	**9.01%**	21.15%	**30.30%**	6.94%	10.53%	3.57%	3.42%	9.62%	11.83%
4	**0.90%**	2.88%	12.12%	**52.78%**	22.37%	19.05%	0.00%	2.88%	7.53%
5	**1.80%**	10.58%	1.52%	16.67%	**44.74%**	13.10%	2.56%	5.77%	3.23%
6	**0.00%**	0.00%	13.64%	18.06%	10.53%	**41.67%**	0.00%	2.88%	26.88%
7	**23.42%**	1.92%	0.00%	0.00%	0.00%	0.00%	**48.72%**	5.77%	2.15%
8	**18.92%**	6.73%	6.06%	0.00%	0.00%	2.38%	13.68%	**29.81%**	12.90%
9	**6.31%**	8.65%	13.64%	0.00%	5.26%	19.05%	0.85%	19.23%	**29.03%**

RF	1	2	3	4	5	6	7	8	9
1	**26.13%**	10.53%	3.17%	0.00%	0.00%	1.33%	27.73%	13.33%	4.12%
2	11.71%	**43.42%**	22.22%	1.41%	0.06%	0.00%	9.24%	9.17%	5.15%
3	9.91%	26.32%	**22.22%**	8.45%	12.64%	0.00%	0.84%	12.50%	15.46%
4	0.90%	2.63%	14.29%	**46.48%**	27.59%	17.33%	0.00%	1.67%	9.28%
5	4.50%	13.16%	14.29%	21.13%	**44.83%**	0.07%	0.84%	4.17%	4.12%
6	0.00%	5.26%	7.94%	19.72%	9.20%	**50.67%**	0.00%	5.83%	17.53%
7	21.62%	1.32%	1.59%	0.00%	0.00%	0.00%	**49.58%**	5.83%	1.03%
8	18.02%	7.89%	3.17%	0.00%	0.00%	2.67%	10.92%	**30.83%**	13.40%
9	7.21%	13.16%	11.11%	0.00%	2.30%	21.33%	0.84%	16.67%	**29.90%**

Model	Accuracy	Kappa
K-Nearest Neighbour	23%	0.14
Gradient Boosting Trees	37%	0.29
Random Forest	40%	0.29

Table 3: *Results of multi class prediction on aggregated sample .*

Results on disaggregated sample

		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
KNN		***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
	1	**73.00%**	2.00%	15.00%	0.00%	3.00%	1.00%	6.00%	10.00%	0.00%
	2	0.00%	**67.00%**	0.00%	3.00%	0.00%	0.00%	1.00%	0.00%	0.00%
	3	0.00%	2.00%	**60.00%**	0.00%	0.00%	0.00%	2.00%	0.00%	2.00%
	4	0.00%	0.00%	0.00%	**97.00%**	0.00%	0.00%	0.00%	0.00%	0.00%
	5	20.00%	0.00%	42.00%	0.00%	**87.00%**	4.00%	4.00%	15.00%	1.00%
	6	2.00%	0.00%	13.00%	0.00%	2.00%	**64.00%**	2.00%	2.00%	8.00%
	7	0.00%	18.00%	17.00%	0.00%	0.00%	0.00%	**67.00%**	1.00%	0.00%
	8	1.00%	10.00%	29.00%	0.00%	1.00%	0.00%	19.00%	**70.00%**	1.00%
	9	3.00%	0.00%	29.00%	0.00%	6.00%	31.00%	0.00%	2.00%	**88.00%**
GBM		***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
	1	**88.49%**	1.23%	2.24%	1.22%	4.36%	0.00%	3.43%	4.14%	0.24%
	2	0.59%	**84.06%**	0.00%	6.97%	0.04%	0.00%	6.63%	1.56%	0.00%
	3	0.32%	1.28%	**81.34%**	1.51%	0.21%	0.00%	1.11%	0.28%	0.96%
	4	0.00%	0.09%	0.00%	**83.52%**	0.00%	0.00%	0.00%	0.00%	0.00%
	5	6.82%	0.32%	5.22%	0.75%	**90.38%**	4.00%	2.18%	5.28%	0.00%
	6	1.26%	0.00%	5.97%	0.00%	0.88%	**92.22%**	0.05%	0.51%	3.37%
	7	0.09%	8.70%	0.75%	4.05%	0.18%	0.00%	**78.35%**	3.15%	0.00%
	8	1.22%	4.33%	2.24%	1.98%	2.14%	0.22%	8.25%	**84.82%**	0.00%
	9	1.22%	0.00%	2.24%	0.00%	1.83%	3.56%	0.00%	0.26%	**95.42%**
RF		***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
	1	**82.73%**	0.96%	9.62%	0.00%	8.27%	0.00%	4.29%	5.00%	0.35%
	2	0.60%	**84.50%**	0.00%	2.64%	0.00%	0.00%	6.92%	2.24%	0.00%
	3	0.60%	1.33%	**75.00%**	0.00%	0.31%	0.27%	2.54%	1.34%	0.70%
	4	0.00%	0.23%	0.00%	**97.14%**	0.00%	0.00%	0.00%	0.00%	0.00%
	5	8.01%	0.23%	3.85%	0.00%	**83.67%**	6.04%	3.66%	7.67%	1.40%
	6	2.26%	0.23%	5.77%	0.00%	2.27%	**70.60%**	0.79%	1.22%	23.16%
	7	0.64%	9.77%	1.92%	0.22%	0.21%	0.00%	**67.97%**	2.76%	0.00%
	8	3.22%	2.75%	3.85%	0.00%	0.31%	0.27%	13.63%	**79.23%**	0.70%
	9	1.93%	0.00%	0.00%	0.00%	4.96%	22.80%	0.20%	0.54%	**73.68%**

Model	Accuracy	Kappa
K-Nearest Neighbour	65%	0.58
Gradient Boosting Trees	80%	0.73
Random Forest	79%	0.75

Table 4: *Results of multi class prediction on disaggregated sample* .

Interim conclusions

- Can predict fuel poverty from consumption behavior.
- Shortening the feedback loops for energy company decision makers (and regulators).
- Improve analysis with linkage of building stock, socio-demographic, and weather data.